

Development of Neurofuzzy Architectures for Electricity Price Forecasting

Abeer Alshejari ¹, Vassilis S. Kodogiannis ^{2,*} and Stavros Leonidis ³

¹ Department of Mathematical Science, Princess Nourah bint Abdulrahman University, PO Box 84428, Riyadh 11671, Saudi Arabia; aaalshejari@pnu.edu.sa

² School of Computer Science & Engineering, University of Westminster, London W1W 6UW, United Kingdom

³ School of Pure and Applied Sciences, Open University of Cyprus, 2220 Nicosia, Cyprus; stavros.leonidis@st.ouc.ac.cy

* Correspondence: V.Kodogiannis@westminster.ac.uk

Received: 30 January 2020; Accepted: 3 March 2020; Published: 5 March 2020

Abstract: In 20th century, many countries have liberalized their electricity market. This power markets liberalization has directed generation companies as well as wholesale buyers to undertake a greater intense risk exposure compared to the old centralized framework. In this framework, electricity price prediction has become crucial for any market player in their decision-making process as well as strategic planning. In this study, a prototype asymmetric-based neuro-fuzzy network (AGFINN) architecture has been implemented for short-term electricity prices forecasting for ISO New England market. AGFINN framework has been designed through two different defuzzification schemes. Fuzzy clustering has been explored as an initial step for defining the fuzzy rules while an asymmetric Gaussian membership function has been utilized in the fuzzification part of the model. Results related to the minimum and maximum electricity prices for ISO New England, emphasize the superiority of the proposed model over well-established learning-based models.

Keywords: day-ahead electricity price forecasting; neurofuzzy systems; neural networks; clustering; prediction

1. Introduction

During the past decades, energy markets have experienced considerable changes. The energy liberalization in many countries dissolved a centralized integrated market and transformed it into a deregulated competitive one. Market participants now include energy generators, distribution companies, traders, and large consumers who interact bilateral or through various exchange trading power spots, like European Power Exchange and Nodal Exchange and PJM in the United States [1,2].

The aim of an electricity market is to reduce the electricity cost through this liberalization framework. From this point of view, electricity could be considered as a “commodity” like other “trading” products in market. However, this specific “commodity” is different from other financial services because of its unique characteristics. As electrical energy cannot be stored, a continuous stable condition between power generation and demand is always required. In addition, this energy “commodity” is a typical day-ahead market that does not permit for continuous trading like other financial markets. This is due to the fact that system operators need in advance information for checking the power schedule is in line with any transmission constraints. In a day-ahead market situation, trading agents on day ($d - 1$) are required to submit their bids/offers for electricity supply for each hour of the next day (i.e., day d) before a specific market closing time [3]. The goal for an energy market is to balance both the electricity demand and the supply and thus determine

eventually the market clearing price (MCP). MCP is considered as the price for the entire system, however in case of transmission congestion issues, a different factor, the locational marginal prices (LMP), has to be defined. The prediction of LMP factor is considered as very significant for market participants, however because of its level of complexity such forecasting is assumed to be more difficult compared to MCP factor [4]. An independent system operator, who acts as a transmission service provider, controls the overall transmission network.

Electricity price forecasting (EPF) is thus an important financial-related process for any decision-making system in power sector. This financial characteristic has initiated substantial research contributions in energy forecasting. A large portion of these contributions is associated with electricity load forecasting and particularly short-term load forecasting (STLF) [5]. However, MCPs/LMPs are considered to have more volatile characteristics than hourly loads and thus EPF prediction is more complex than the STLF. MCP/LMP is dependent on a number of features, such as historical hourly price and load data, system operating conditions (which include any possible predicted power shortfalls and generation outages), weather conditions, time indices (including hours, week days, holidays, and multiple seasonality), bidding strategies, and operating reserves/imports [6]. Even though electricity price curves retain similar changes with load curves in a transaction day, they depend on more uncertain factors. As a result, prediction error for the MCP/LMP is generally greater than the hourly load under similar conditions [7]. The difficulty in MCP/LMP modelling is also influenced by another important factor. Because of the recent changes in power market, there is a growing interest for the use of renewable energy sources (RES). At the same time, market is interested in those technologies that provide flexibility and dispatch-ability to the power grid, in other words technologies to be able to balance changes in power supply/demand in cases when RES are not offered. Coal and gas-fired power plants are considered as flexible technologies, thus an additional important factor is the difference between the market electricity price and its cost of production. For gas-fired cases, this parameter is called spark spread, while for coal-fired assets, dark spread. The importance of spark spread forward curve is high to energy industry planners, as it helps them to address generation profits [8].

EPF models generally can be classified based on their horizon duration, mainly in three groups, namely, short-term price forecasting (STPF), medium-term price forecasting (MTPF), and finally long-term price forecasting (LTPF) schemes. STPF approach includes next hour and day-ahead price prediction schemes and is important for a quick decision-making process, so that producers and consumers can schedule their bidding strategies. MTPF includes prediction for next week, next month, up to one year. The performance of this specific scheme is affected by seasonal effects, such as rise in electricity price in summer because of higher load consumption and decline in winter. MTPF information can be used by suppliers to optimize their production cost by planning an efficient resource allocation for generation of electricity. Finally, LTPF horizon varies from couple of year to decades. Such models are used by policy-makers to plan pricing schemes and management of resources, while investors utilize them for analyzing recovery of investment in power plant construction, production, type of energy sources, and transmission.

The importance of EPF as well as its complexity level, especially for the STPF case has motivated research, and several data-driven approaches have been proposed by researchers. Following the restructuring process in energy sector, time-varying autoregressive (AR) models have been explored in analyzing price time-series in Japan Electric Power Exchange spot market. The estimation of AR's coefficients was performed via Kalman Filter algorithm [9]. A seasonal autoregressive integrated moving average (ARIMA) model has been also used for the prediction of day-ahead electricity prices in Elspot market in Sweden, which is considered as one of the main energy trading "day-ahead" market in the world [10]. Additionally, other time-series models, utilizing dynamic regression methodologies have been proposed for electricity price prediction for PJM market [11]. An interesting feature of this approach is the adoption of a clustering scheme as a way to minimize the resulting model complexity. While an ARIMA model is suitable for forecasting applications, its main weakness is that it is not able to handle efficiently high levels of nonlinearity and volatility that exist in electricity price time-series. A combination of generalized autoregressive

conditional heteroskedastic (GARCH) models and ARMA has been successfully demonstrated for predicting the electricity price in ISO New England market [12]. Alternatively, the combination of ARIMA models and signal processing techniques, such as wavelets, has been explored for the Singapore electricity market [13]. Although some promising results have been achieved using these linear techniques, they cannot model adequately the non-linear characteristics of the electricity price. In cases where high-frequency changes of the electricity price occur, the application of these methods could be considered as problematic [14]. In recent years, alternative methodologies based on computational intelligent techniques, including neural networks (NN) have been investigated by researchers for EPF. Multilayer perceptrons (MLP) and radial basis function (RBF) NNs have been successfully used in forecasting the electricity price volatility in PJM market, the United States [15].

Electricity price signal characteristics, such as non-stationarity and time variance, have been addressed with the aid of appropriate feature selection techniques [16]. Similarly, MLP-based models have been explored to predict next-week prices in California as well as Spanish electricity markets [17]. These NN approaches outperformed the ARIMA as well as the naïve methods used in all cases. This reveals the superiority of NN-based systems over methods which have been traditionally used in energy forecasting systems. An Elman recurrent NN has been also utilized to EPF for the mainland Spain market, using samples captured during winter and summer weeks, achieving a superior to MLP-based NNs performance [18]. The further development of NNs has led to the evolution of deep learning (DL)-based systems. Long-short term memory (LSTM) and convolutional neural network (CNN) based models have been applied and compared against ARIMA and single hidden-layer MLP models for the day-ahead price market in Belgium. Results revealed a clear advantage of DL models over traditional approaches; however their performance against MLP networks was rather comparable [19].

As an alternative methodology to NNs, support vector machines (SVMs) and probabilistic-based techniques have grown in popularity in recent years. SVMs are not prone to local minima problems like MLPs, as they provide a nonlinear mapping of the original data into a higher dimension space. This desired characteristic has been addressed in research papers associated to the electricity load and price area [20]. Generally, NNs and SVM methods focus on determining the predicted electricity price for a specific given horizon. However, in some cases, such information is not enough to provide an efficient risk analysis. Thus, instead of predicting the future electricity price, there is an interest in estimating the uncertainty involved in that specific price, knowing the degree of spreading of the forecasted value and their probabilities [21]. Such uncertainty in price is influenced by the level of demand, which in turn depends on the intensity of business activities and weather. With the incorporation of renewable energies into electric grid, the effect of uncertainty has been growing significantly, especially with the wind power generation, where in periods of high generation, the electricity price is likely to be lower [22]. A methodology for probabilistic forecasting, where a Bayesian stochastic volatility model with double exponential distribution of spikes and exogenous variables for the PJM market has been recently explored [23].

One of the fundamental problems for efficient system identification using MLP, which is a well-known NN, is the lack of neurons' multi-dimensional "characteristics," as the multi-dimensional hypersphere of the input domain for each neuron node cannot be defined through the use of sigmoid activation function. On the other hand, the use of bell-shaped activation function, like the Gaussian function, includes such "multi-dimensional" characteristics, as its center parameter could hold n "dimensions" equal to the number of input variables. This is also explained by the local vs. global learning styles involved in different learning-based architectures. Fuzzy logic systems alternatively can offer such input space association; nevertheless, no learning ability is taking part for rules definition. Hybrid intelligent systems provide some modelling advantages and an adaptive-network-based fuzzy inference system (ANFIS) neurofuzzy (NF) model has been implemented for electricity price forecasting [24]. Such models proved to have superior performance against related simulation studies that utilize ARIMA and MLP schemes [14,18]. Based on the

existing research studies in STPF, it seems that the investigation of hybrid neurofuzzy (NF) models needs further consideration.

Obviously, the area of EPF is too complex and a number of parameters should be considered for an efficient forecasting process. These include three areas: Pre-processing analysis of data signals, feature selection of input variables and finally, the design of the prediction model. Pre-processing analysis involves consideration of any trend/seasonality/outliers (spikes), stationarity issues, and decomposition mode techniques [25]. Feature selection is a rather complicated task, as the number and the type of input variables need to be carefully chosen. This includes historical values of prices as well as a number of exogenous factors such as electricity load, day type identification codes (i.e., weekdays/holidays) temperature, fuel prices. The implementation of an efficient prediction model and its associated issues (type of model—MIMO, MISO, recurrent) complete the whole design of the forecasting framework. The main objective of this paper is aiming at supplementing research in the area of NF systems and their application in hourly ahead EPF. In this specific research, as a proof of concept, we focus mainly on the prediction model and more precisely on the development of novel clustering-based NF modelling schemes and their application in the hourly step ahead EPF. The equally important issues of pre-processing analysis with the incorporation of decomposition mode techniques and the inclusion of additional exogenous parameters, in addition to electricity load, will be considered in a future work.

In general, two main forecasting approaches have been explored for the hourly ahead EPF. In the former case, one MIMO model is used as a 24-hourly prices predictor, while in the latter case a recurrent NN configuration has been investigated. However, the MIMO approach incorporates, by nature, a network structure with relative high complexity as it has to “accommodate” a large number of input variables to model adequately all 24-hourly next-step-ahead prices [26,27]. With this type of complexity, such MIMO model may result under-fitting forecasts for specific hours, while in the same time-release over-fitting forecasts for different hours. This problem is explained by the fact that it is a complicated process and a difficult task to simultaneously model all the dynamics of 24-hourly prices via a single MIMO scheme. Alternatively, although a recurrent model could result in a less complex structure, in practice the feedback error accumulation could deteriorate its performance. In order to avoid such intrinsic problems in recurrent NNs, one option is to configure the input vector of a recurrent NN with a nonlinear autoregressive exogenous structure (NARX); however the overall complexity of the network will be increased dramatically [28].

Alternatively, a modular structure with 24 MISO models, where each hour is modelled via a dedicated modelling learning-based unit, has been adopted in this paper for EPF [29]. As each hourly-model is modelled independently, its complexity, including the number of input variables is much reduced. The proposed scheme is different from existing approaches and in this paper, as a pilot study, we have implemented and demonstrated two MISO models of real data obtained from New England, USA system operator associated only to 22:00 h and 04:00 h which correspond to the hours where electricity prices have the maximum and minimum values respectively. The proposed novel NF scheme utilized as a one-step-ahead prediction model is the asymmetric-based Gaussian Fuzzy Inference Neural Network (AGFINN) architecture. This specific architecture has been designed by taking into consideration the following sub-objectives:

- A NF model should have a small as possible structure, i.e., rules.
- A NF model could be generic in terms of architecture, acting either as MISO or MIMO.
- A NF model may need to have a different nonlinear membership function.

Clustering analysis could be utilized for the definition of the number of fuzzy rules as well the estimation of the initial membership function parameters in the premise/fuzzification part. In the proposed AGFINN design, this task has been assigned to the Fuzzy C-means (FCM) method. One of the most successful, in terms of applicability, NF architectures is the Takagi-Sugeno-Kang (TSK), with ANFIS scheme as its main representative [30]. However, ANFIS’s performance is deteriorated in cases of large number of fuzzy rules. In the proposed AGFINN model, the main problem of “curse of dimensionality” is considerably reduced, as the numbers of membership functions (MF) for each input variable and fuzzy rules are equal. Additionally, AGFINN uses an asymmetric MF rather the

standard symmetric Gaussian MF usually employed in NF systems. It has been considered from literature, that variability and flexibility features are higher in asymmetric Gaussian functions compared to the standard one [31]. Thus, the main innovation in this proposed architecture is the development of a clustering-based non-symmetric based NF system. Although the concept of non-symmetric MF exists already in literature, it is much more complicated and requires more parameters to be tuned compared to the approach proposed in this paper [32]. The RLS learning training algorithm utilized in ANFIS model has been also adopted here, as AGFINN shares with ANFIS also a TSK defuzzification layer. In this research study, AGFINN has been implemented acting either as an MISO or MIMO system. In the MISO configuration, a Takagi-Sugeno based defuzzification scheme has been adopted, while for the MIMO case, a “center average” (CA) scheme has been utilized as defuzzifier. The proposed models are compared against classic ARIMA models as well as adaptive fuzzy logic systems (AFLS) [33], ANFIS, and MLP models in order to evaluate their forecasting performance. The comparison against alternative NF models is considered essential as these specific methodologies have proved to provide robust performances for EPF based on literature review.

2. The Asymmetric Neurofuzzy Architecture

The proposed AGFINN architecture is considered as a novel neurofuzzy modelling scheme. The first proposed AGFINN design is illustrated by a traditional MIMO structure, shown in Figure 1, which includes the FCM method as a clustering initialization step. The gradient descent training algorithm has been used as a learning scheme, while AGFINN’s output is calculated via a “center average” (CA) defuzzification method. Many NF models, however, are following the TSK-based style, which has the advantage of modelling nonlinear processes with relatively fast training speed, because of the presence of linear units in the defuzzification part. Hence, a TSK-based scheme has been also developed to investigate the possible advances over ANFIS. The proposed AGFINN-TSK has a structure and architecture similar to AGFINN-CA, as shown in Figure 2.

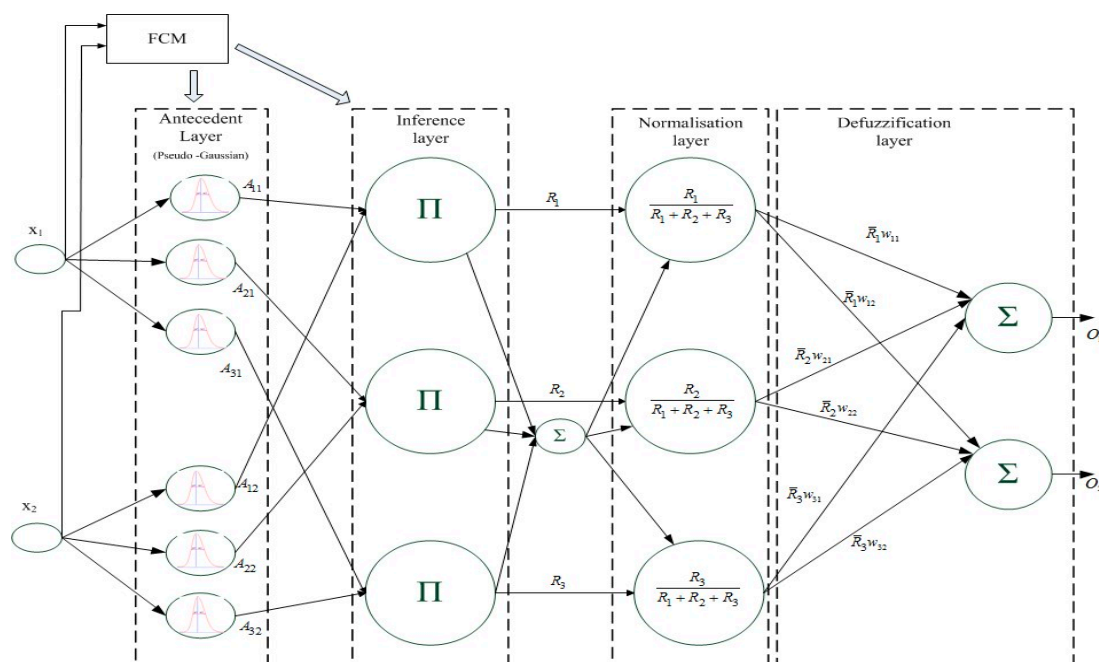


Figure 1. Illustration of AGFINN-CA system.

The first three layers L_1 , L_2 , and L_3 are associated to the premise part (i.e., IF part) of a fuzzy system, while layer L_5 relates to the equivalent consequent part (i.e., THEN part). Layer L_4 represents the normalization layer, and is applied to the outputs from L_3 . In this paper, centers derived from the fuzzy c-means clustering method were used for the initialization for the centers of fuzzy MFs. In the

fuzzification part of AGFINN model, the number of fuzzy rules equals also with the number of clusters. In addition, the spread parameters for each MF σ_{ij} are initialized based on the equation 1, where elements in matrix \mathbf{U} correspond to the MFs of input x_k from the i^{th} cluster, as they were derived from the FCM clustering scheme used as a preprocessing step.

$$\sigma_{ij} = \left(\sum_{k=1}^n u_{ik} (x_{kj} - c_{ij})^2 / \sum_{k=1}^n u_{ik} \right)^{1/2} \quad (1)$$

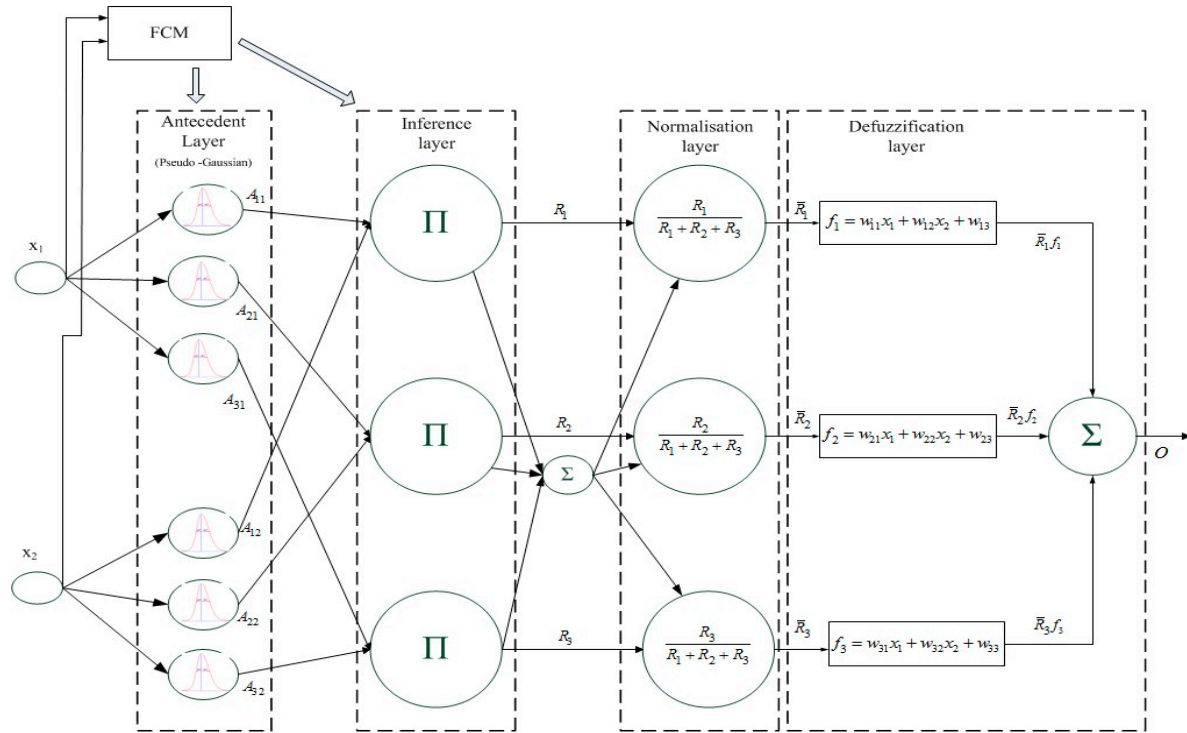


Figure 2. Illustration of AGFINN-TSK system.

2.1. AGFINN: Forward Learning Analysis

The FCM pre-processing step practically indicates the generation of the fuzzy rules base. Thus, fuzzy rules can be formulated by the following equations, based on the adopted configuration:

$$\text{IF } (x_1 \text{ is } U_1^i \text{ AND} \dots \text{AND } x_q \text{ is } U_q^i) \text{ THEN } \begin{cases} y_k = \sum_{j=1}^c w_{kj} R_j, & (\text{MIMO-CA}) \\ y = \sum_{j=1}^c f_j R_j, \quad f_j = w_{j1}x_1 + \dots + w_{jn}x_n + w_{j(n+1)} & (\text{MISO}) \end{cases} \quad (2)$$

where U are the sets derived from the c -partition of training data X and R_c are the fuzzy normalized rules. The proposed configuration of AGFINN schemes is thus illustrated as:

- Layer 1: Nodes at this stage simply forward input variables x_1, x_2, \dots, x_n to L2.
- Layer 2: This is premise part in a fuzzy IF-THEN structure, where utilizes an asymmetric Gaussian MF with the following form

$$A_{ij} = \exp \left(-\frac{1}{2} \left(\frac{x_i - c_{ij}}{\sigma_{ij}^{left}} \right)^2 \right) U(x_i; -\infty, c_{ij}) + \exp \left(-\frac{1}{2} \left(\frac{x_i - c_{ij}}{\sigma_{ij}^{right}} \right)^2 \right) U(x_i; c_{ij}, \infty) \quad (3)$$

$$\text{where } U(x_i; a, b) = \begin{cases} 1 & \text{if } a \leq x_i < b \\ 0 & \text{otherwise} \end{cases}$$

The MF in Equation (3) includes two types of spreads, namely σ_{ij}^{left} and σ_{ij}^{right} respectively, which transform the classic Gaussian MF to a non-symmetric output form, as shown in Figure 3. Moreover, this approach utilizes a “pseudo” function that requires two spreads—left and right—and a common center parameter, instead of the Gaussian interval type-2 asymmetric membership function (IT2MFs) that requires two distinct symmetric functions with individual parameters [32]. The adopted approach not only is less complicated but also provides a truly asymmetric form for MFs.

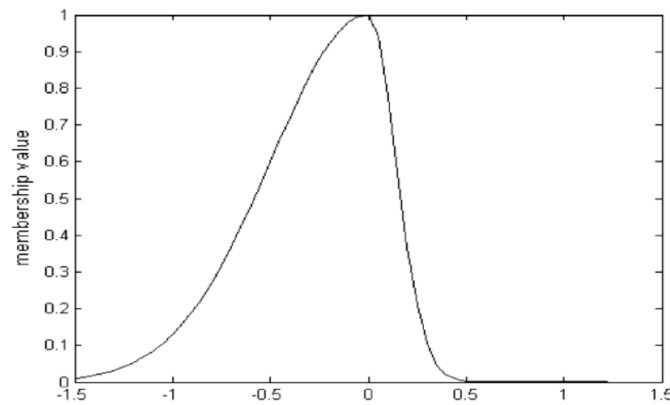


Figure 3. Structure of asymmetric membership functions (MF).

Following the FCM clustering phase, an initial value for spread (σ_{ij}^{init}) has been calculated. As AGFINN utilizes two spreads, one located at the left of the initial center parameter and one at the right, both spreads are initialized as $\sigma_{ij}^{init} / 2 = \sigma_{ij}^{left} = \sigma_{ij}^{right}$. Thus, during the first iteration of the training process, the total spread of any asymmetric MF has been equaled to $\sigma_{ij}^{total} = \sigma_{ij}^{left} + \sigma_{ij}^{right}$. Upon the arrival of any input variable from L_1 , its position (left/right) against the specific center parameter for each MF needs to be recorded via a specific MF index allocated for each MF. This index is then used in the backward learning phase to update that particular spread parameter. The value of this index is then updated accordingly to any new input arrival from L_1 . During forward training phase, σ_{ij}^{total} is used as the spread used in the Gaussian function which has the specific form

$$A_{ij} = \exp \left(-\frac{(x_i - c_{ij})^2}{2b_{ij}^2} \right) \quad (4)$$

where $b_{ij} = \sigma_{ij}^{total}$, i represent the number of MF/rules, while j denotes the specific input variable. During the backward learning phase, a new spread σ_{ij}^{new} value is calculated via the GD learning method. Based on the information stored at that specific MF index, either the left or right spread is

updated as $\sigma_{ij}^{left \text{ or } right} = \sigma_{ij}^{new} / 2$. For the next iteration step, in the forward training phase, the spread parameter will be equal again as $\sigma_{ij}^{total} = \sigma_{ij}^{left} + \sigma_{ij}^{right}$, incorporating however the relative adjustment of one of its components.

- Layer 3: This layer represents the “classic” rules layer. The multiplication has been used as a fuzzy AND operation, thus output has the following form:

$$R_i = \prod_j^n A_{ji}(x_j) \quad (5)$$

In this proposed architecture, the number of rules is the same as the number of clusters.

Layer 4: At this normalization layer, each normalized rule is calculated by:

$$\bar{R}_i = \frac{R_i}{\sum_{j=1}^c R_j} \quad (6)$$

- Layer 5: Finally, this is the consequent stage of the AGFINN scheme. Based on the specific configuration, either CA or TSK, AGFINN’s output has the following form

$$\begin{cases} O_i = \sum_{j=1}^c w_{ij} \bar{R}_j, & (AGFINN - CA) \\ O = \sum_{j=1}^c f_j \bar{R}_j, & (AGFINN - TSK) \end{cases} \quad (7)$$

where $f_j = w_{j1}x_1 + \dots + w_{jn}x_n + w_{j(n+1)}$ illustrates the “consequent linear parameters” of the TSK-based scheme.

2.2. AGFINN: Backward learning analysis

AGFINN’s training involves the usage of the gradient descent (GD) algorithm for the parameters’ tuning for AGFINN-CA scheme. For the case the AGFINN-TSK scheme, a hybrid training algorithm, which includes the use of RLS method for the consequent parameters, has been implemented. During, this phase, correction signals are calculated from the AGFINN’s output backward to the premise part of the model and all network’s parameters are adjusted. During training phase, for each p^{th} training data pair (x_p, y_p) , the parameters are adjusted in order to minimize the following objective function

$$E = \frac{1}{2} \sum_{k=1}^K (D_k - O_k)^2 \quad (8)$$

where K is the number of outputs and D_k the desired response of the k^{th} output. Variable O_k is defined as in Equation (7). Based on GD method, the weights of the AGFINN-CA scheme at the defuzzification layer are updated through the equation

$$\Delta W_{ki} = -\frac{\partial E}{\partial W_{ki}} = -\frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial W_{ki}} = (D_k - O_k) \bar{R}_i \quad (9)$$

where $k=1$ and $i=1, 2, \dots, c$ denote the number of output and normalization units respectively. Thus, the weights at the defuzzification layers are updated as:

$$W_{ki}(t+1) = W_{ki}(t) + \eta_w \Delta W_{ki} \quad (10)$$

where η_w is the learning rate for training. The c_{ij} and b_{ij} parameters of the proposed membership function are tuned via the following equation

$$\begin{aligned}
 c_{ij}(t+1) &= c_{ij}(t) - \eta_c \left(\frac{\partial E}{\partial c_{ij}} \right) \\
 b_{ij}(t+1) &= b_{ij}(t) - \eta_b \left(\frac{\partial E}{\partial b_{ij}} \right)
 \end{aligned}
 \tag{11}$$

where $\frac{\partial E}{\partial c_{ij}}$ and $\frac{\partial E}{\partial b_{ij}}$ are the components extracted from the following chain rule configuration:

$$\begin{aligned}
 \frac{\partial E}{\partial c_{ij}} &= \frac{\partial E}{\partial O} \frac{\partial O}{\partial R_i} \frac{\partial R_i}{\partial A_{ij}} \frac{\partial A_{ij}}{\partial c_{ij}} \\
 \frac{\partial E}{\partial b_{ij}} &= \frac{\partial E}{\partial O} \frac{\partial O}{\partial R_i} \frac{\partial R_i}{\partial A_{ij}} \frac{\partial A_{ij}}{\partial b_{ij}}
 \end{aligned}
 \tag{12}$$

All learning rates parameters (η_w, η_c, η_b) in our simulation have been set with a constant value of 0.15. The partial derivative components in Equation (12) are then calculated as

$$\begin{aligned}
 \frac{\partial A_{ij}}{\partial c_{ij}} &= \left[\exp \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right) \right]' = \left[\exp \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right) \right] \cdot \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right)' = \\
 &= A_{ij} \cdot \left(-\frac{1}{2b_{ij}^2} \left((x - c_{ij})^2 \right)' \right) = A_{ij} \cdot \left(\frac{1}{b_{ij}^2} (x - c_{ij}) \right)
 \end{aligned}
 \tag{13}$$

$$\begin{aligned}
 \frac{\partial A_{ij}}{\partial b_{ij}} &= \left[\exp \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right) \right]' = \left[\exp \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right) \right] \cdot \left(-\frac{(x - c_{ij})^2}{2b_{ij}^2} \right)' = \\
 &= A_{ij} \cdot \left(-\frac{(x - c_{ij})^2}{2} \left(\frac{1}{b_{ij}^2} \right)' \right) = A_{ij} \cdot \left(\frac{(x - c_{ij})^2}{b_{ij}^3} \right)
 \end{aligned}
 \tag{14}$$

$$\frac{\partial R_i}{\partial A_{ij}} = \prod_{lm \neq ij} A_{lm}
 \tag{15}$$

$$\frac{\partial O}{\partial R_i} = \frac{w_i \left(\sum_{\substack{j=1 \\ j \neq i}}^c R_j \right) - \sum_{\substack{j=1 \\ j \neq i}}^c R_j w_j}{\left(\sum_{j=1}^c R_j \right)^2}
 \tag{16}$$

Inspired by ANFIS's hybrid learning algorithm, similar analysis has been performed for the case of AGFINN-TSK configuration. In nonlinear models that utilize least-squares (LS) methods, system's output can be expressed through a regression model as $Y = \Phi^T \theta$, where Φ is the co-called regressor and θ represents the parameter vector. The basic principle in LS estimation is to calculate the optimal values for parameters that minimize the square of the difference between the estimated and the real output of a nonlinear function. For small datasets, the "batch" version of LS could be easily applied; however, as the number of sample increases, computations can become computationally too expensive because of the calculation of the inverse of $\Phi^T \Phi$. RLS method can

provide an efficient alternative solution to update the θ estimate parameter vectors each time a new sample appears, without using all the previous samples in the calculation and without computing the inverse of $\Phi^T \Phi$ [28]. The system output of the AGFINN-TSK scheme, shown in Equation (7), is re-formulated then as:

$$(O = \sum_{l=1}^c \sum_{j=1}^{(n+1)} ((\bar{R}_l Z_j) w_{lj})) \quad (17)$$

where l is the number of nodes at L_4 , and Z denote the number of input variables plus 1 (i.e., $Z = [x_1, x_2, \dots, x_n, 1]$). The consequent linear parameters are denoted as w_{ij} . Let us denote

$$(\Phi = [\bar{R}_1 x_1, \bar{R}_1 x_2, \dots, \bar{R}_1 x_n, \bar{R}_1, \dots, \bar{R}_c x_1, \bar{R}_c x_2, \dots, \bar{R}_c x_n, \bar{R}_c]) \quad (18)$$

$$\theta = [w_{11}, w_{12}, \dots, w_{1n}, w_{1(n+1)}, \dots, w_{c1}, w_{c2}, \dots, w_{cn}, w_{c(n+1)}] \quad (19)$$

the vector parameters to be utilized in the RLS estimation method. With a new input sample $\Phi(t)$, and desired output value $d(t)$. RLS procedure is summarized as:

- Update the input history vector $\Phi(t)$
- Compute the model output using the previous set of model linear coefficients $\theta(t-1)$

$$y(t) = \Phi^T(t) \theta(t-1) \quad (20)$$

- Compute the error

$$e(t) = d(t) - y(t) \quad (21)$$

- Compute the gain vector

$$K(t) = P(t) \Phi(t) \quad (22)$$

Where

$$P(t) = \frac{1}{\lambda} \left[P(t-1) - \frac{P(t-1) \Phi(t) \Phi^T(t) P(t-1)}{\lambda + \Phi^T(t) P(t-1) \Phi(t)} \right] \quad (23)$$

- Update the linear coefficients for the next iteration

$$\theta(t) = \theta(t-1) + K(t) e(t) \quad (24)$$

The value of weighting factor λ is considered as system's memory, and is related with the convergence and the ability of the method to track time-varying statistics in the input sequence. In this research, this parameter has been set constant to one. Initialization of $P(0)$ has been set as $P(0) = aI$, for some large positive number a . All modelling schemes have been implemented in MATLAB (ver. R2016a, Mathworks.com).

3. The Short-Term Electricity Price Forecasting

The performance of the proposed AGFINN models are examined on the hourly day-ahead electricity market of New England, USA (ISONE—Independent System Operator of New England), commonly used as a benchmark case study in related literature [34]. ISO New England (ISONE)

(<http://www.iso-ne.com>) is a not-for-profit independent corporation responsible for managing the daily operation of the power grid in New England, USA.

In order to improve the system's reliability and to mitigate prices volatility, ISONE manages day-ahead and forward capacity markets by scheduling future supply and demand needs. ISONE is the power system regulator for the six-state region of New England and serves 7.1 million retail electricity customers in a population of 14.7 million. In total, ISONE has control over 350 electricity generators (coal, oil, nuclear power, natural gas, and renewable energy) within the region and maintains energy ties with neighboring regional grids in New York, Quebec, and New Brunswick. Existing power units have approximately 30,500 MWs of installed electricity generating capacity that covers the non-constant power demand in the region with high electricity peak of 28,000 MWs in the summer period, and smaller peak of 22,800 MW in the winter.

In this research, ISONE data from the period 2006–2007 was used for the development of the proposed prediction models. Initially, data were divided into 24 hourly time-series and for each hourly time-series the first 600 data were used for training purposes, while the remaining 123 ones for testing. Based on the proposed forecasting framework, each one of these 24 MISO models is scheduled to provide a next-day forecast for a specific hour. In this paper, only prediction results associated with the maximum (22:00h) as well as the minimum (04:00h) electricity price are shown. Similarly, the same process can be expanded for each hour. The specific New England price time-series used in this research are shown in Figures 4 and 5 respectively. The specific choice is justified by the fact that the maximum price has the highest average value among remaining hours, while its plot of the raw data reveals nonlinear characteristics with arbitrary amplitude. On the other hand, minimum price reveals a rather smoother response, however with the presence of frequent spikes.

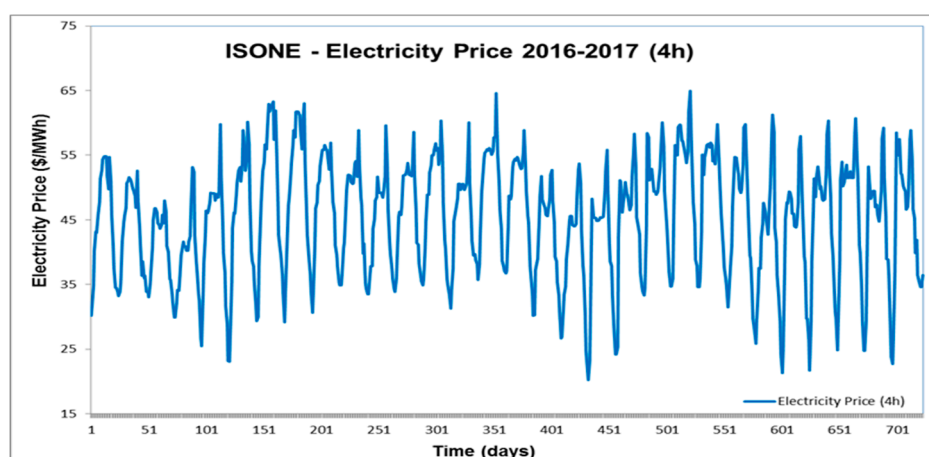


Figure 4. ISO-NE electricity price times—series corresponding to 4 h.

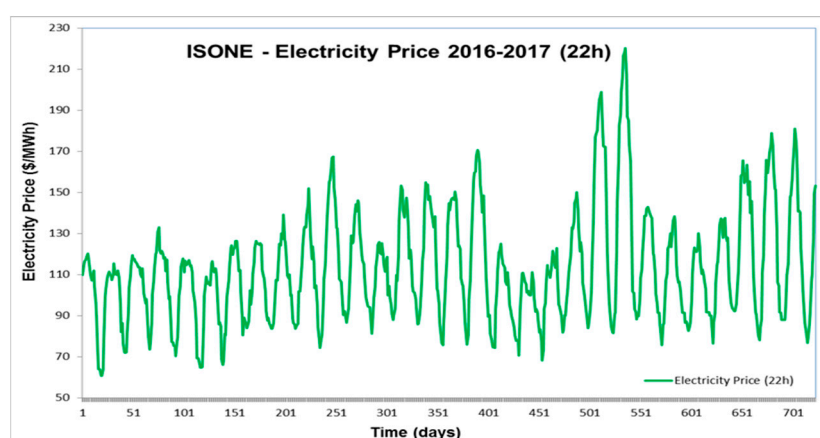


Figure 5. ISO-NE electricity price times—series corresponding to 22 h.

In this research study, as the primary aim is to explore the efficiency of AGFINN models against other learning-based approaches, analysis was performed only on raw data provided by ISONE. It has to be added that the presence of nonlinear characteristics in electricity price curves appears generally in any benchmark datasets and it is independent of any specific time period [35]. Obviously, pre-processing analysis (incl. seasonality, trend, stationarity issues and decomposition mode methods) as well as selection of appropriate input variables will be considered in a future work, in order to provide a complete assessment of the applicability of neurofuzzy systems in this area. An efficient forecasting methodology in electricity prices depends heavily on the specific input parameters used in the system identification process. In general, internal historical price values and external factors such as temperature and/or electricity load as well as day-type (i.e., weekends, holidays) are considered as potential input variables. This problem therefore involves the consideration of appropriate feature selection techniques for choosing the suitable input variables. In this research study, as an initial pilot study, the well-known Boruta feature selection method has been employed to explore the level of importance of adopted input variables [36]. Based on the existing literature, electricity load has been assumed as an important exogenous factor [37]. Hence, in this research, only next day's predicted electricity load together with historical electricity prices have been considered as potential input variables. Therefore, in order to investigate further the performance of AGFINN models, three different case studies have been implemented in this research. As it is very interesting to determine whether there is any significant difference between the performances of AGFINN models in these studies, the Diebold-Mariano test [38] has been used to explore performances obtained from AGFINN models.

4. Results and Discussion

In order to evaluate the performance of AGFINN NF models, all studies were carried out based on the same New England electricity price datasets. For the evaluation of the forecasting performance of the all developed models, various error measures have been employed as assessment criteria. These include root mean square error (RMSE), absolute percentage error (APE), mean absolute percent error (MAPE), mean absolute error (MAE), standard error of prediction (SEP), and the coefficient of determination (R^2) [37]. Relative measures are another type of evaluation indices, with Theil U_1 coefficient as one of the most used measure [37]. The Theil coefficient is scale invariant and its range is between one and zero. If U_1 equals zero then a perfect fit is assumed. The main idea of utilizing such relative measures is to assess the performance of a forecast relative to a benchmark forecast [39]. Measures may result in very big values because of either outliers or poor modelling, which in turn make the comparison of different forecasts not feasible or even not consistent. Relative measures could remove the bias introduced by possible trends, outliers or even seasonal components, provided that the benchmark forecast handles these issues appropriately.

Initially, an approach utilizing ARIMA models has been attempted. ARIMA scheme, a well-known forecasting method, consists of three components, the autoregressive (AR), integral, and the moving average (MA) part. AR explores the use of the time-delayed inputs to predict the future value. Integral component is related to differencing required to transform the data stationary, while MA is similar to AR, as it is used to predict the error using historical time-delayed errors [40]. The fact that ARIMA models are able to predict the target variable only based on its historical data, make them very powerful. Especially, for short run forecast of hourly or daily data, they can provide with high level of accuracy small ups and downs. On the other hand, in cases where the target variable increased or decreased dramatically and unexpectedly, ARIMA may provide much less accurate predictions.

ARIMA model's parameters are associated to the number of past (time-delayed) input variables incorporated in the model (i.e., p), to the number of times that inputs are differenced (i.e., d) and finally to the size of the moving average window (i.e., q). In this study ARIMA models have been created using the XLSTAT software (v.2019.2, Addinsoft) for both 4 h and 22 h cases. After some trials using classical time series ARIMA, two models with two input variables have been developed in both case studies. More specifically, for the 4-h case, an ARIMA(3,1,0) has been

implemented while an ARIMA(4,1,0) was constructed for the 22 h respectively. Testing results are shown in Figure 6, while the values for RMSE, SEP, and MAPE indices for the 4-h case were 5.79, 12.96%, and 10.90 respectively, while the corresponding values for 22h were 8.67, 7.17%, and 5.45. Based on the graphical illustrations in Figure 6, both curves clearly reveal to have different characteristics, with the 22-h case to retain a much smoother output compared to the 4-h case. However, as the main aim of this paper is to propose new NF architectures, existing NF schemes such as AFLS [33] and ANFIS [30] were implemented and tested under the same conditions. In addition, a comparison with an MLP NN has been explored, as this specific network has been widely used in the past alongside the classic ARIMA models. Our approach toward the MLP network however is different, as our MLP models were implemented with a two-hidden layer structure instead of one layer. The standard gradient descent algorithm (GD) has been adopted as the MLP's learning method. It is well-known that historical information of electricity prices and load demand are considered as important inputs for forecasting the electricity price. In this research, the proposed forecasting architecture has been based on an hourly horizon plan. In all presented models, only historical prices and forecasted demands have been chosen as appropriate input variables, while three case studies based on different input selection schemes have been considered for all used learning-based methods. The choice of this input configuration, as showing following case studies, has been influenced from previous related studies with NF and NN models respectively [37,41]. A series of trials has been carried out in order to define AGFINN's optimal configuration in terms of internal structure, number of clusters, etc.

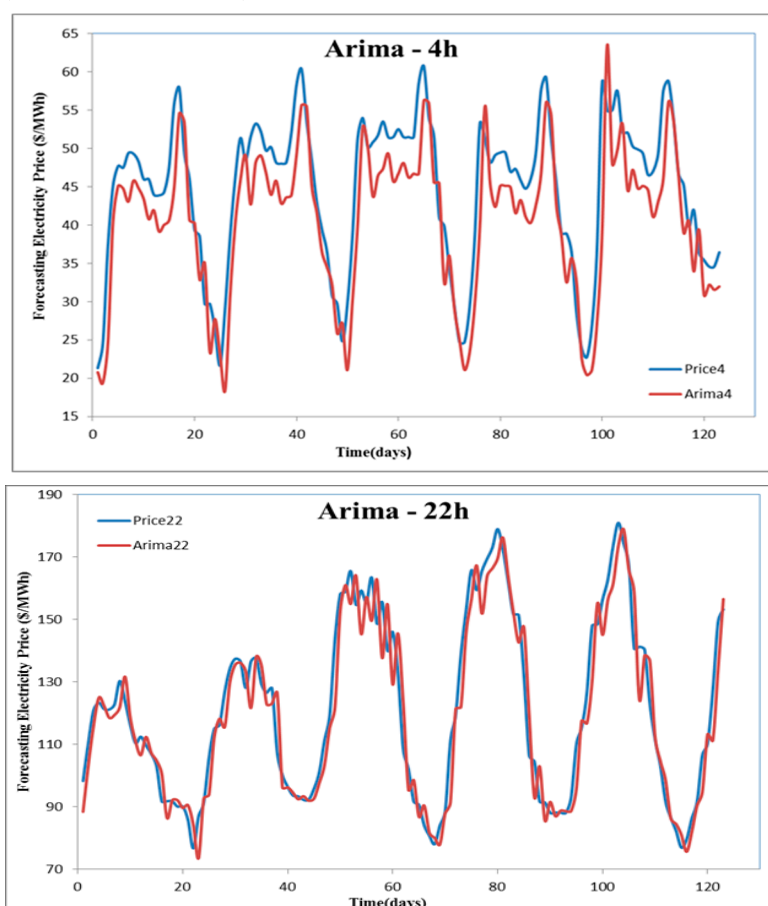


Figure 6. Performances of ARIMA models for 4 h & 22 h.

4.1. Case Study 1

In this specific scenario, the forecasted price depends not only on the historical prices (days and hours), but also on the predicted electricity load for that targeted hour/day. Hence, the proposed modeling configuration for a specific hour (i) and day (j) is

$$Price(i, j) = f \left\langle \begin{matrix} Price(i, j-1), Price(i, j-2), Price(i-1, j-1), \\ Price(i-2, j-1), Load(i, j) \end{matrix} \right\rangle$$

where in $Price(i, j)$, i is related to the specific hour, while j to the specific day.

This specific scenario has been considered based on the assumption that of the price ($i, j-1$) position as the main variable is known, while additional variables were selected from its direct “neighborhood.” More specifically, prices on the same hour (i) at two previous days and prices on the same day ($j-1$) at two previous hours have been considered.

In this case study, additionally it is considered that the forecasted demand $Load(i, j)$ is also available via a dedicated “load” forecasting model. In a complete energy forecasting framework, an additional model for the prediction of load demand is also required. Such “load” model will provide the predicted $Load(i, j)$ information required by the “price” model. Obviously, the predicted load input variable will introduce an additional error to the overall error generated by the “price” model. However, as it is necessary to associate input $Load(i, j)$ with output $Price(i, j)$, it is of great importance that a high accurate “load” model needs to be developed as well. AGFINN models have been utilized in prediction of the maximum (22 h) and minimum (04 h) price, using this specific input configuration. Best results were produced by AGFINN-RLS models.

After trying various internal structures, the optimal one utilized 20 and 15 rules for the cases of 22 h and 04 h respectively. The combination of GD and RLS algorithms as learning methods resulted in a fast training process, which was concluded in less than 500 epochs. Performances for these two specific hours, based on testing datasets, are shown in Figures 7 and 8 for all simulated forecasting models, while Tables 1 and 2 summarize their statistical performances.

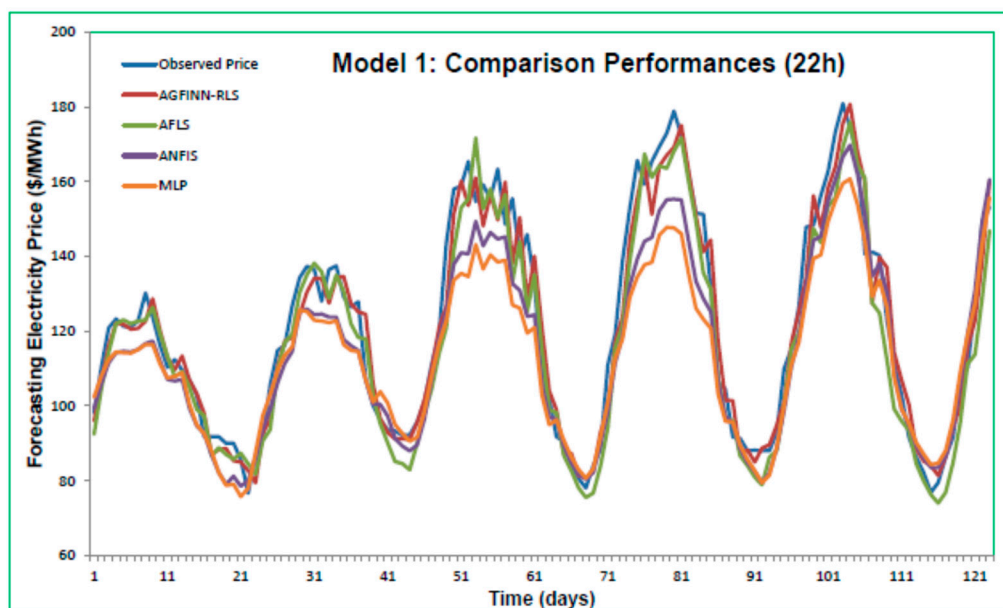


Figure 7. Prediction for electricity price at 22:00—case study 1.

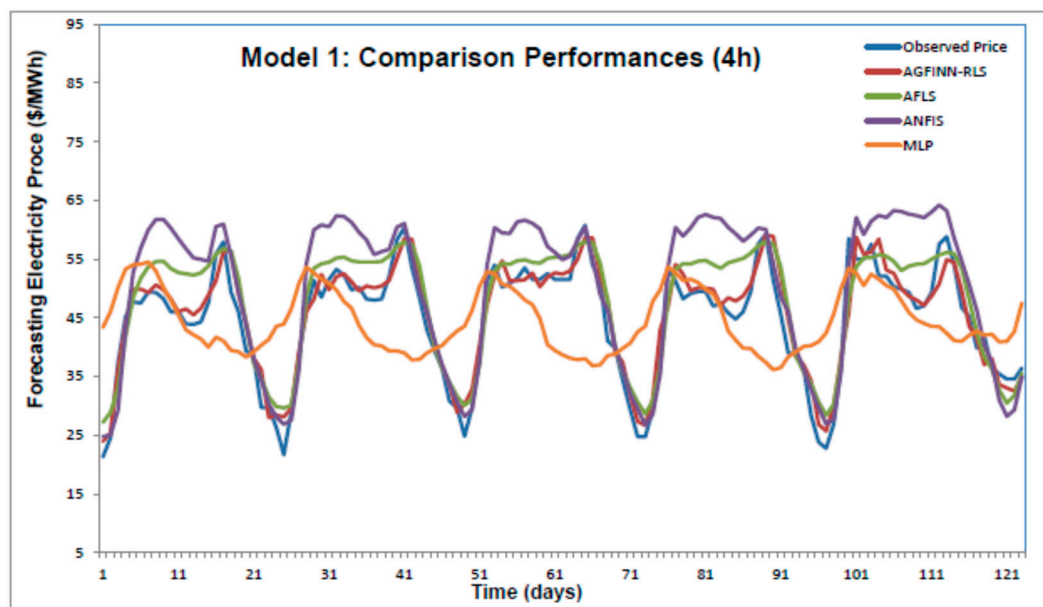


Figure 8. Forecasting for electricity price at 04:00—case study 1.

Table 1. Statistical performance for electricity price forecasting models at 22:00—case study 1.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	7.2511	8.1100	8.6060	10.3569	12.8880
APE	551.0082	607.4207	671.4143	775.0515	912.2255
MAPE	4.4797	4.9384	5.4587	6.3012	7.4165
MAE	5.5696	6.3175	6.7172	8.1925	9.8865
SEP	5.9924	6.7022	7.1122	8.5591	10.6508
U_1	0.0293	0.0332	0.0353	0.0430	0.0539
R^2	0.9679	0.9729	0.9664	0.9694	0.9619

A close inspection of these curves reveals that AGFINN-RLS model fitting with historical price data managed to forecast the result very satisfactory.

Table 2. Statistical performance for electricity price forecasting models at 04:00—case study 1.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	3.1753	4.0542	4.6194	7.5882	10.4905
APE	727.0175	956.1771	1169.1	1675.5	2627.2
MAPE	5.9107	7.7738	9.5050	13.6217	21.3592
MAE	2.4390	3.3526	3.9456	6.1493	8.1517
SEP	7.1039	9.0700	10.3345	16.9763	23.4693
U_1	0.0344	0.0433	0.0490	0.0779	0.1158
R^2	0.9516	0.9482	0.9325	0.9147	0.1321

Although the “observed price” curves in both figures have different dynamic characteristics, it is significant to indicate the existence of some “volatile” price spikes appeared in both cases. From the statistical tables, the clear advantages of AGFINN-based models against other models implemented for this specific problem are clearly shown. Case Study 1 can be used as a benchmark for all implemented models for the next two case studies, where alternative input variable schemes will be presented. In fact, AGFINN-CA’s performance could be considered to be close to AFLS’s one,

while the superior performance of TSK-based model is due to their specific defuzzification scheme. Although AGFINN models performed very satisfactory in both cases (22 h and 4 h), they need to be compared against other NF and neural network models, in order to extract more useful conclusions.

One such competitor is the adaptive fuzzy logic system (AFLS) architecture. AFLS shares like AGFINN the CA defuzzifier output as well as the AGFINN's premise part, without however the pre-processing clustering stage. A classic Gaussian function has been used in its MFs. After trying different number of MFs/Rules and the GD method as learning scheme, eventually 25 and 20 fuzzy rules were chosen as the optimal ones for the cases of 22 h and 04 h respectively. Additionally, MLP and ANFIS models have been developed for this case study. Both structures have a great range of applicability, thus such comparison is considered as a benchmark. After a few trials, an MLP neural network with 20 and 8 nodes employed at the first and second hidden layer respectively while the adopted ANFIS NF scheme utilized 32 rules. The proposed AGFINN architecture has certain structural advantages over the ANFIS model, as the number of its MFs in fuzzification layer is equal to the numbers of rules. The increased number of MFs increases the localization of the input signal while simultaneously retains low the number of fuzzy rules. The main problem with ANFIS design is the excessive number of fuzzy rules which is a result of combination of input variables/number of MFs. AGFINN-RLS scheme shares the same hybrid learning algorithm as in the case of ANFIS.

4.2. Case Study 2

It has been investigated through research literature, that there is high level correlation of the current hourly electricity price with related prices at two and seven days before respectively [41].

This issue implies the existence of a daily and weekly periodicity. Hence, the aim in this case study is to explore this hypothesis.

No external variables (i.e., load) have been considered as potential input variables. The proposed modeling configuration for a specific hour (i) and day (j) is

$$Price(i, j) = f \left\langle \begin{matrix} Price(i, j-1), Price(i, j-2), Price(i, j-3), \\ Price(i, j-7), Price(i-1, j-1), Price(i-2, j-1) \end{matrix} \right\rangle$$

In this proposed case study, the number of input variables was increased to six, as weekly information of the electricity price for that specific hour was also considered. Performances for these two specific hours, based on testing datasets, are shown in Figures 9 and 10 for all simulated forecasting models, while Tables 3 and 4 summarize their statistical performances.

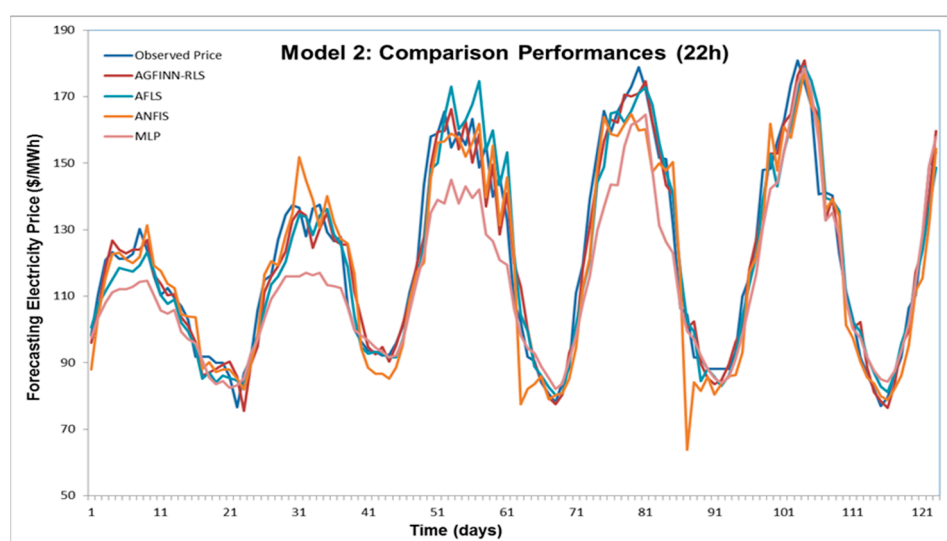


Figure 9. Prediction for electricity price at 22:00—case study 2.

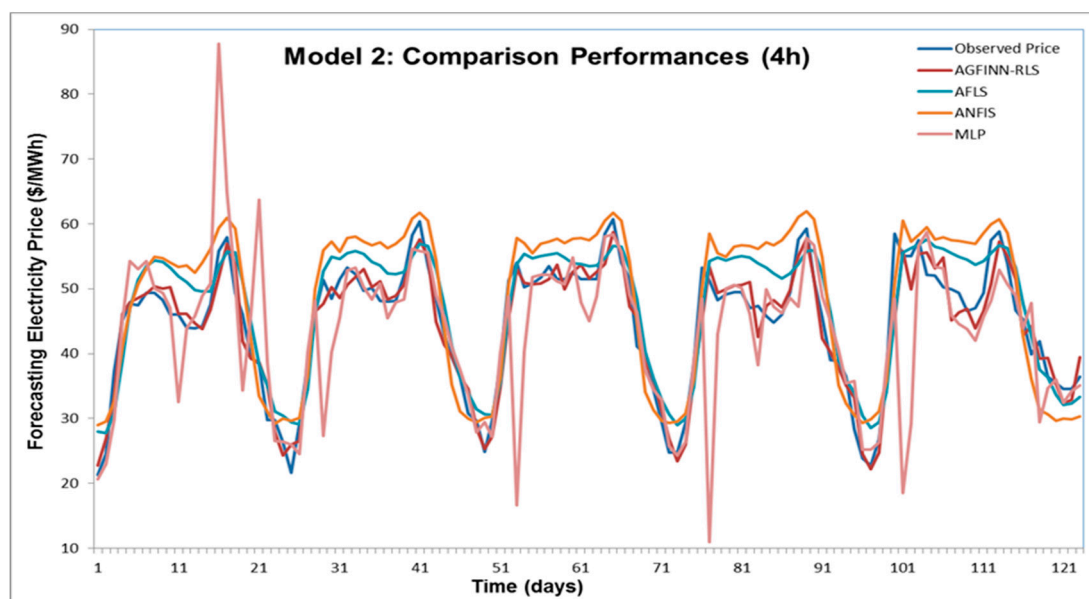


Figure 10. Forecasting for electricity price at 04:00—case study 2.

Best results were produced again by AGFINN-RLS models. From both forecasting graphs and statistical performances, an improvement in results accuracy, compared to case study 1, has been noted. The proposed hypothesis in this case study, based on the produced improved results, was verified, thus the characteristics of periodicity could be included for case study 3.

Table 3. Statistical performances for electricity price forecasting models at 22:00—case study 2.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	6.9356	7.6515	7.9278	9.8380	11.6525
APE	540.905	560.288	591.165	759.612	855.803
MAPE	4.3976	4.5552	4.8062	6.1757	6.9578
MAE	5.2993	5.7602	5.9638	7.4261	9.0974
SEP	5.7317	6.3233	6.5517	8.1302	9.6298
U_1	0.0280	0.0311	0.0320	0.0400	0.0484
R^2	0.9703	0.9660	0.9620	0.9462	0.9552

The optimal structure for AGFINN-RLS included 20 fuzzy rules for both cases. As it is illustrated from these statistical tables, AFLS models utilizing 20 and 25 rules for 4 h and 22 h respectively, achieved a satisfactory performance, although inferior to AGFINN-CA model. Both AFLS and AGFINN-CA shared the same type of defuzzification scheme. Both MLP and ANFIS forecasting results, in this case study, were problematic as they revealed a number of unexpected “spikes” and 4 h and 22 h respectively. The MLP retained the same network configuration as in case study 1, while in the case of ANFIS, a larger in size network with 64 fuzzy rules has been created, increasing thus the computational cost.

Table 4. Statistical performances for electricity price forecasting models at 04:00—case study 2.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	2.9220	3.7748	4.5427	6.0034	8.7796
APE	637.484	929.999	1160.1	1515.4	1306.4
MAPE	5.1828	7.5610	9.4320	12.3199	10.6214
MAE	2.2178	3.1964	3.9305	5.2979	4.8608
SEP	6.5371	8.4450	10.1630	13.4307	19.6418
U ₁	0.0321	0.0402	0.0484	0.0629	0.0969
R ²	0.9555	0.9580	0.9214	0.9197	0.6799

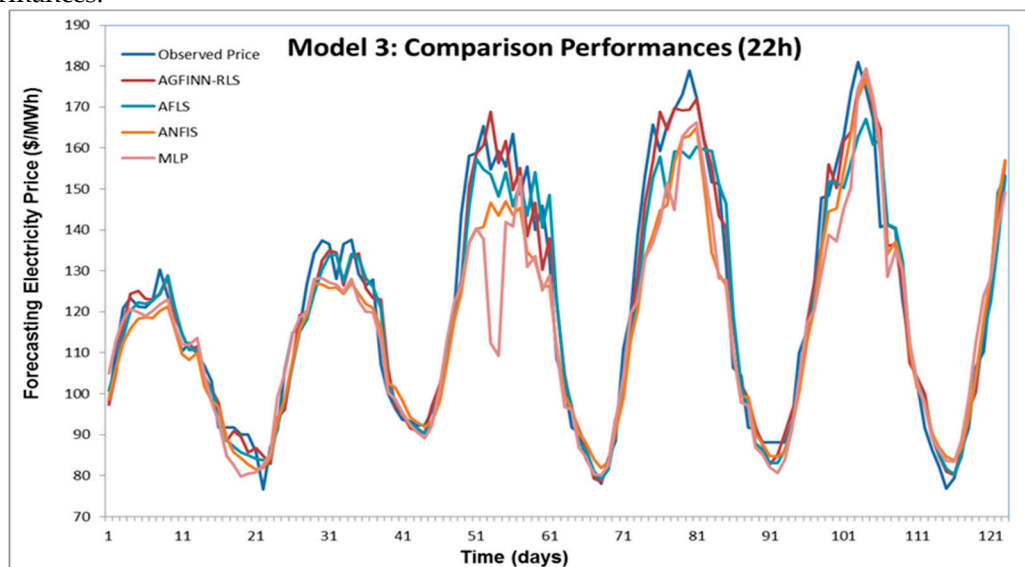
4.3. Case Study 3

The objective of this last case study was to expand the case study 2 by incorporating external input variables such as predicted electricity load from case study 1. It was considered as very important to evaluate the combined effect of exogenous input together with the weekly periodicity characteristics of the price. The proposed modeling configuration for a specific hour (i) and day (j) is

$$Price(i, j) = f \left\langle \begin{matrix} Price(i, j-1), Price(i, j-2), Price(i, j-3), Load(i, j) \\ Price(i, j-7), Price(i-1, j-1), Price(i-2, j-1) \end{matrix} \right\rangle$$

In this proposed case study, the number of input variables was increased to seven because of the inclusion of the forecasted value of electricity load.

Performances for these two specific hours, based on the testing datasets, are shown in Figures 11 and 12 for all simulated forecasting models, while Tables 5 and 6 summarize their statistical performances.

**Figure 11.** Prediction for electricity price at 22:00—case study 3.

As it is illustrated from Figure 11, AGFINN and AFLS models managed to capture the dynamics (i.e., trend) at 4:00 h with different levels of accuracy. For this experiment, AGFINN-RLS utilized 20 rules, while the optimal number of rules for AFLS was 25. On the other hand, forecasting via ANFIS model resulted in unexpected spikes. This issue is due to the complexity of the specific ANFIS model. With seven input variables and two MFs per input, a massive number of 128 rules required for training purposes.

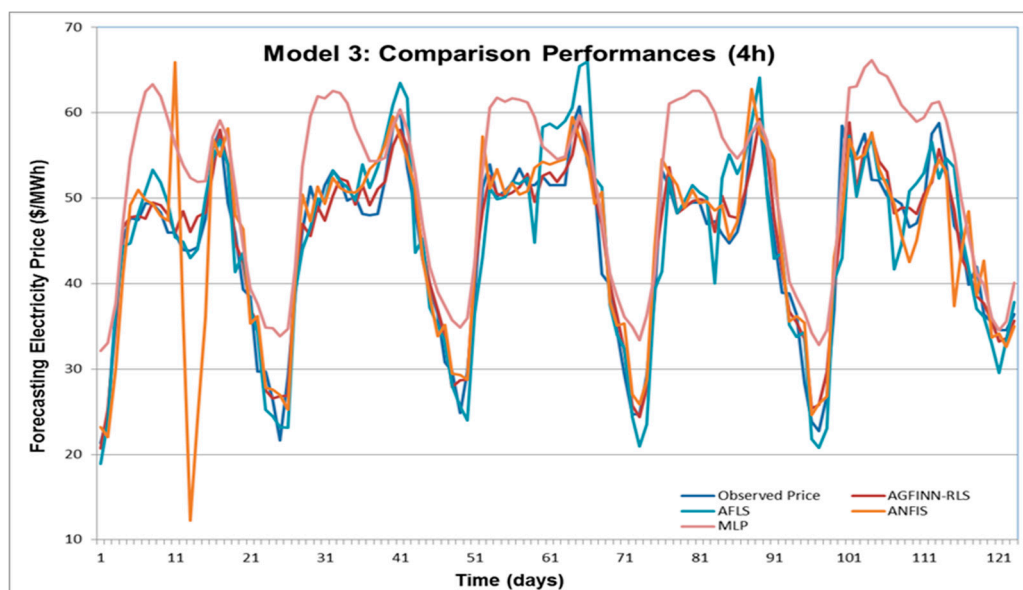


Figure 12. Prediction for electricity price at 04:00—case study 3.

Table 5. Statistical performances for electricity price forecasting models at 22:00—case study 3.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	6.4605	7.5032	7.7340	9.1584	11.4835
APE	487.999	616.196	550.734	655.688	739.416
MAPE	3.9675	5.0097	4.4775	5.3308	6.0115
MAE	4.8768	5.9290	5.7301	6.9329	7.9360
SEP	5.3391	6.2007	6.3915	7.5686	9.4901
U ₁	0.0260	0.0303	0.0314	0.0377	0.0473
R ²	0.9742	0.9699	0.9670	0.9682	0.9388

ANFIS's performance was significantly improved for the case of 22 h, as shown in Figure 12. MLP's performance was deteriorated however for this case. AGFINN models in both experiments outperformed "rival" models, as shown also from the related statistical tables. Although, case study 3 proved to be the most accurate compared to the previous cases, even in this case, there were variations in performance between the two AGFINN-based models.

Table 6. Statistical performances for electricity price forecasting models at 04:00—case study 3.

Stats	AGFINN-RLS	AGFINN-CA	AFLS	ANFIS	MLP
RMSE	2.6612	3.5844	4.3667	5.4409	8.0055
APE	608.757	857.814	944.293	1023	2040.3
MAPE	4.9492	6.9741	7.6772	8.3168	16.5878
MAE	2.0810	2.9241	3.34409	3.5089	6.8644
SEP	5.9536	8.0189	9.7692	12.1724	17.9098
U ₁	0.0290	0.0383	0.0474	0.0590	0.0815
R ²	0.9635	0.9588	0.9229	0.8630	0.9079

All models utilize FCM clustering units, as well as asymmetric functions as MFs. The difference in defuzzification part is mainly responsible for such variations. It was verified that TSK-based models are more accurate/suitable for forecasting applications compared to CA-based systems.

Although this specific case proved to provide superior results compared to the other cases, it is also important to assess the performance of all these learning-based methods against the ARIMA regression scheme.

Results revealed that for both electricity price datasets, ARIMA models performed quite satisfactorily and outperformed the performance of MLP neural networks. Such results are comparable with existing studies in literature, and can be explained by the relatively smooth (without any abrupt spikes) characteristics of both hourly datasets.

Overall, the electricity price forecasting problem is a difficult task that involves the implementation of many related subtasks, such as pre-processing analysis, efficient input variables selection and accurate “price” forecasting models either for one or multi steps-ahead predictions. In addition, the implementation of an accurate “load/demand” prediction model would benefit the forecasting performance of the “price” model. In this research study, the performance of AGFINN model against existing learning-based models, justified the concept of its development. However, there are a number of open questions related to the problem of accurate electricity price forecasting. Prediction models need to be incorporated in the whole framework of analysis, which means that issues of pre-processing and input variables selection need to be addressed. Although these issues will be considered in a future research, at this stage it is important to assess the performances of RLS-based AGFINN models as they have been applied in these three case studies. Practically, model 1 included information related to the predicted load, while model 2 included information of the previous weekly electricity price. Model 3 incorporated both of them. Figure 13 illustrate the performances of RLS-based AGFINN models for these three case studies for the case of 22 h.

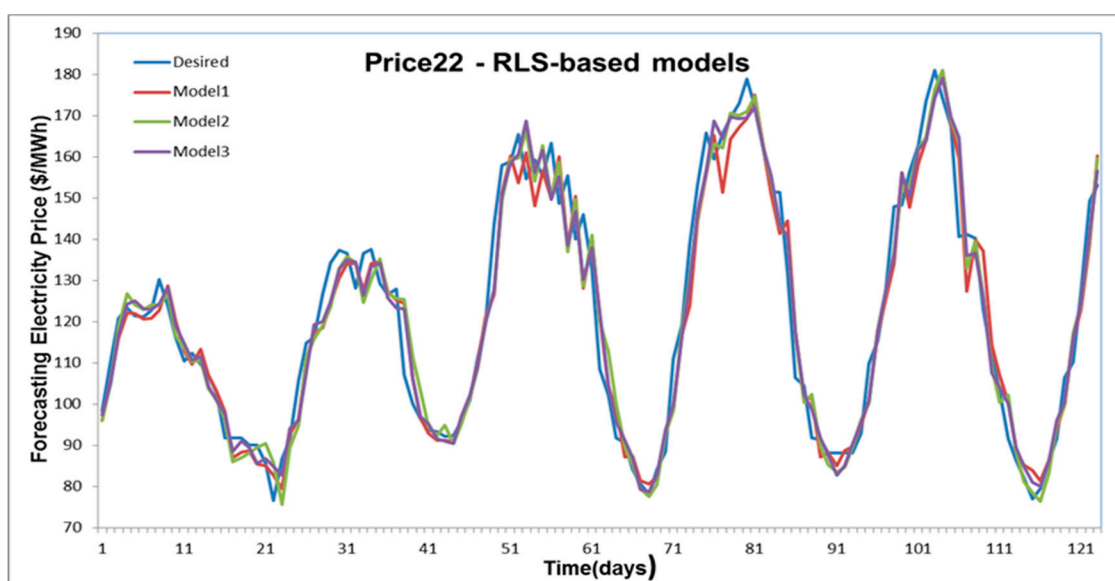


Figure 13. Performance of RLS-based models for electricity price at 22:00.

Tables 1, 3, and 5 summarize their performances based on established indices, however two issues need to be discussed: How different are these three RLS based models and which of the chosen input variables are more important.

To test the statistical significance of the results shown in Figure 13, the Diebold-Mariano (DM) test, which is the state-of-the-art method to evaluate significance was implemented using Excel. Furthermore, the DM test is used to show the difference of these three RLS-based AGFINN models through the calculation of the related p -values for the DM test. The main objective is to explore whether this difference shown is significant (for predictive purposes) or simply because of the specific choice of data values in the sample [42]. By comparing model 1 and model 2, the p -value was equal to $0.36 > 0.05$, thus no significant difference exists between the two forecasts. The comparison of model 2 and model 3 resulted in a p -value equal to 0.0915, which was slightly above the 0.05 threshold. However, the comparison of model 1 with model 3 resulted in a p -value equal to

0.012 which was less than the 0.05 threshold, indicating thus a significant difference. In fact, the difference between these two models was the incorporation of the past weekly price information in model 3, which is an indication that this variable could be considered as an important factor.

Therefore, in order to identify the level of importance of those seven input variables employed in the case study 3, the Boruta method has been utilized [43,44]. Boruta is an ensemble-based feature selection algorithm that utilizes random forest (RF) principles with additional mechanisms for finding better results. This technique follows an all-relevant feature selection process where it captures all features which are in some conditions related to the desired output. In contrast to Boruta, other “classic” feature selections algorithms follow a minimal optimal process where they depend on a small set of features which results in a minimal error on a selected predictive model. This specific feature selection technique has been built in an R-language environment (v.3.6.2), and results are illustrated through a graphical boxplot graph. Red and green boxplots represent Boruta output-scores of rejected and confirmed features respectively, while blue boxplots correspond to minimal, average, and maximum output-score of a shadow feature.

Dataset corresponding to prices at 04:00 h and 22:00 h have been processed through this mechanism, and results are shown in Figures 14 and 15.

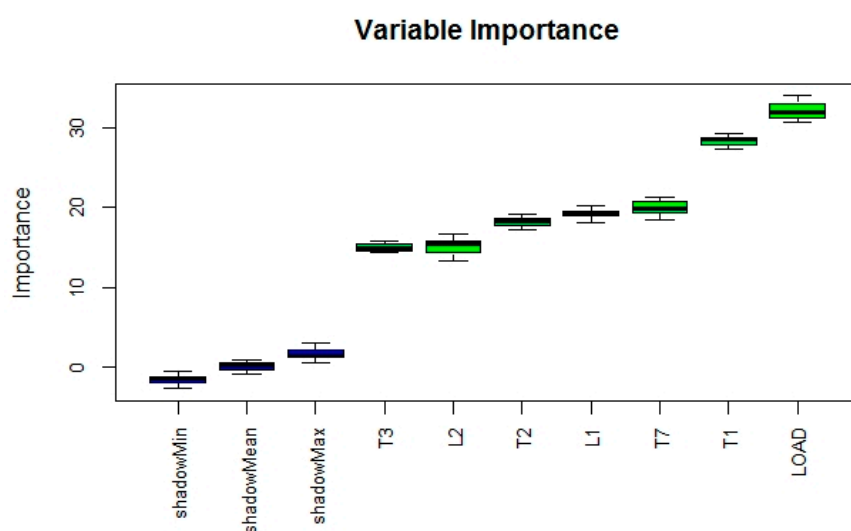


Figure 14. Boruta analysis for electricity price at 04:00.

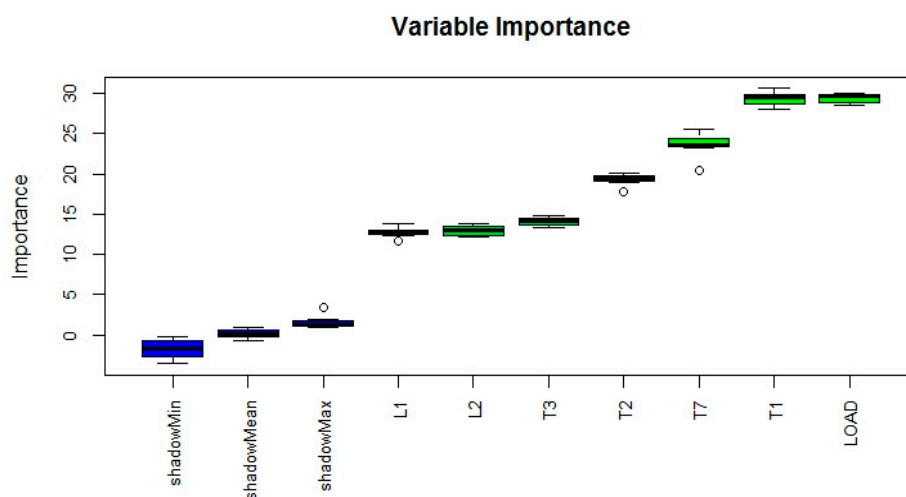


Figure 15. Boruta analysis for electricity price at 22:00.

A common characteristic in both figures is that although all chosen input variables were considered as appropriate ones, the load factor is considered as the one with the highest level of importance. This is another justification of the close relationship between load and price. Obviously, the current electricity price is considered an important factor (i.e., 2nd factor in both boxplots) however the 7th past price variable has been clearly defined as the 3rd most important factor. This indicates that the hypothesis of the second case study was correct.

5. Conclusions

Currently, energy liberalization in many countries forced both power companies and regulators to develop efficient electricity price forecasting models. This paper presented an hourly ahead short-term electricity price forecast by using a novel Neurofuzzy modelling approach using information from ISO New England market. The proposed AGFINN model utilized the FCM methodology identifying the number of MFs/rules, while incorporates asymmetric MFs in the fuzzification section. Two alternative approaches have been investigated based on the defuzzification choice. The effectiveness of the proposed approaches has evaluated through the comparison with neurofuzzy and neural network models. The first case study investigated a simple configuration where historical electricity prices as well as forecasted electricity load were used as the input variables. The objective of the second case study was to consider additional prices at two and seven previous days with no exogenous input (i.e., load) variables. The third case study was considered as a combination of the previous cases, where the effect of using both exogenous inputs together with the weekly periodicity characteristics of the price was evaluated. This specific case proved to be the most accurate, revealing however, how complex the electricity price problem forecasting is in terms of defining appropriate input variables. In future work, the proposed AGFINN prediction model will be integrated with suitable pre-processing and features selection processes in order to provide a complete analysis of energy forecasting problems. Overall, the obtained results from this study revealed the efficiency of neurofuzzy models against classic MLP neural network and ARIMA models currently utilized in this domain. Although both AFLS and MLP models utilize the same GD method as learning algorithm, the different “philosophy” in designing the NF architecture, via Gaussian MFs, allowed AFLS to achieve higher accuracy. The increased number of MFs increases in parallel the localization of the input signal while simultaneously keeps the number of fuzzy rules at low level. Such localization spread through the MFs, is one significant advantage of NF models against the MLP structure. There is a need to investigate further the deployment of advanced computational intelligent methods in the area of electricity price forecasting, and this paper has contributed toward this goal.

Author Contributions: This paper is a collaborative work of all authors. Conceptualization, V.S.K.; Methodology, A.A. and V.S.K.; Data Curation, A.A. and V.S.K.; Software, A.A., V.S.K. and S.L; Validation, A.A., and V.S.K.; Writing—review and Editing, A.A. and V.S.K.; Visualization, A.A., V.S.K. and S.L; Supervision, V.K.

Funding: This research received no external funding

Acknowledgments: This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

Conflicts of Interest: The authors declare no conflict of interest

Nomenclature

AFLS: Adaptive Fuzzy Logic System

AGFINN: Asymmetric Gaussian Fuzzy Inference Neural Network

ANFIS: Adaptive Neural Fuzzy Inference system

EPF: Electricity Price Forecasting

FCM: Fuzzy C-Means

GD: Gradient Descent

ISO: Independent System Operator

MIMO: Multiple Input Multiple Outputs

MISO: Multiple Input Single Output

MLP: Multi-layer Perceptron

NF: neuro-fuzzy

RLS: Recursive Least Squares

STPF: Short term Price Forecasting

TSK: Takagi-Sugeno-Kang

References

1. Shahidehpour, M.; Yamin, H.; Li, Z. *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Assessment*; Wiley: Hoboken, NJ, USA, 2002.
2. Ahmadi, A.; Charwand, M.; Aghaei, J. Risk-constrained optimal strategy for retailer forward contract portfolio. *Int. J. Electr. Power Energy Syst.* **2013**, *53*, 704–713.
3. Mathur, S.; Arya, A.; Dubey, M. Optimal bidding strategy for price takers and customers in a competitive electricity market. *Cogent Eng.* **2017**, *4*, 1358545.
4. Weron, R. Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future. *Int. J. Forecast.* **2014**, *30*, 1030–1081.
5. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; Wiley: Hoboken, NJ, USA, 2006.
6. Singhal, D.; Swarup, K.S. Electricity price forecasting using artificial neural networks. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 550–555.
7. Amjady, N.; Hemmati, M. Energy price forecasting-problems and proposals for such predictions. *IEEE Power Energy Mag.* **2006**, *4*, 20–29.
8. Karakatsani, N.V.; Bunn, D.W. Intra-day and regime-switching dynamics in electricity price formation. *Energy Econ.* **2008**, *30*, 1776–1797.
9. Ofuji, K.; Kanemoto, S. Price forecasting of Japan electric power exchange using time-varying AR model. In Proceedings of the ISAP2007 International Conference on Intelligent Systems Applications to Power Systems, Niigata, Japan, 5–8 November 2007; p. 4441658.
10. Xie, M.; Sandels, C.; Zhu, K.; Nordström, L. A seasonal ARIMA model with exogenous variables for Elspot electricity prices in Sweden. In Proceedings of the 10th International Conference on the European Energy Market, EEM 2013, Stockholm, Sweden, 27–31 May 2013; p. 6607293.
11. Bandyopadhyay, A.; Roy, S.; Ghosh, D. Forecasting day-ahead price of electricity-A dynamic regression approach. *Int. J. Bus. Excell.* **2013**, *6*, 584–604.
12. Liu, H.; Shi, J. Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Econ.* **2013**, *37*, 152–166.
13. Carpio, K.J.E.; Go, A.M.L.; Roncal, C.K.M. Forecasting day-ahead electricity prices of Singapore through ARIMA and wavelet-ARIMA. *DLSU Bus. Econ. Rev.* **2012**, *22*, 97–118.
14. Amjady, N. Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network. *IEEE Trans. Power Syst.* **2006**, *21*, 887–896.
15. Lin, W.-M.; Gow, H.-J.; Tsai, M.-T. An enhanced radial basis function network for short-term electricity price forecasting. *Appl. Energy* **2010**, *87*, 3226–3234.
16. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Convers. Manag.* **2009**, *50*, 2976–2982.
17. Catalao, J.P.; Mariano, S.J.; Mendes, V.M.; Ferreira, L.A. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electr. Power Syst. Res.* **2007**, *77*, 1297–1304.
18. Anbazhagan, S.; Kumarappan, N. Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network. *IEEE Syst. J.* **2013**, *7*, 866–872.
19. Lago, J.; De Ridder, F.; De Schutter, B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Appl. Energy* **2018**, *221*, 386–405.

20. Mohamed, A.; El-Hawary, M.E. Mid-term electricity price forecasting using SVM. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering, Vancouver, BC, Canada, 15–18 May 2016.
21. Maciejowska, K.; Nowotarski, J.; Weron, R. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *Int. J. Forecast.* **2016**, *32*, 957–965.
22. Zhang, Y.; Wang, J. K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting. *Int. J. Forecast.* **2016**, *32*, 1074–1080.
23. Kostrzewski, M.; Kostrzewska, J. Probabilistic electricity price forecasting with Bayesian stochastic volatility models. *Energy Econ.* **2019**, *80*, 610–620.
24. Catalao, J.P.; Pousinho, H.M.; Mendes, V.M. Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Trans Power Syst.* **2011**, *26*, 137–144.
25. Zahid, M.; Ahmed, F.; Javaid, N.; Abbasi, R.A.; Kazmi, H.S.Z.; Javaid, A.; Bilal, M.; Akbar, M.; Ilahi, M. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. *Electronics* **2019**, *8*, 122.
26. Zheng, J.; Xu, C.; Zhang, Z.; Li, X. Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network. In Proceedings of the 51st Annual Conference on Information Sciences and Systems, CISS 2017, Baltimore, MD, USA, 22–24 March 2017, p. 7926112.
27. Taieb, S.B.; Sorjamaa, A.; Bontempi, G. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing* **2010**, *73*, 1950–1957.
28. Nelles, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*; Springer: Berlin/Heidelberg, Germany, 2001.
29. Kodogiannis, V.S.; Amina, M.; Petrounias, I. A clustering-based fuzzy wavelet neural network model for short-term load forecasting. *Int. J. Neural Syst.* **2013**, *23*, 1350024.
30. Jang, J.S.; Sun, R.C.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice-Hall: Upper Saddle River, NJ, USA, 1997.
31. Rojas, I.; Pomares, H.; Fernandez, F.J. A new methodology to obtain fuzzy systems autonomously from training data. *IEEE Conf. Fuzzy Syst.* **1999**, *1*, 527–532.
32. Zhao, T.; Li, P.; Cao, J. Self-organising interval type-2 fuzzy neural network with asymmetric membership functions and its application. *Soft Computing*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 23, pp. 7215–7228.
33. Kodogiannis, V.; Pachidis, T.; Kontogianni, E. An intelligent based decision support system for the detection of meat spoilage. *Eng. Appl. Artif. Intell.* **2014**, *34*, 23–36.
34. Sahay, K.B.; Tripathi, M.M. Day Ahead Hourly Load and Price Forecast in ISO New England Market using ANN. In Proceedings of the 2013 Annual IEEE India Conference (INDICON), Mumbai, India, 3–15 December 2013.
35. Mosbah, H.; El-Hawary, M.E. Hourly Electricity Price Forecasting for the Next Month Using Multilayer Neural Network. *Can. J. Electr. Comput. Eng.* **2016**, *39*, 283–291.
36. Kursu, M.B.; Rudnicki, W.R. Feature selection with the boruta package. *J. Stat. Softw.* **2015**, *36*, 1–13.
37. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead electricity price forecasting via the application of artificial neural network-based models. *Appl. Energy* **2016**, *172*, 132–151.
38. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
39. Bliemel, F. Theil's Forecast Accuracy Coefficient: A Clarification. *J. Mark. Res.* **1973**, *10*, 444–446.
40. Gao, G.; Lo, K.; Fan, L. Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Mark. *Energy Power Eng.* **2017**, *9*, 120–126.
41. Zhou, H.; Wu, X.H.; Li, G. An ANFIS model of Electricity Price Forecasting Based on Subtractive Clustering. In Proceedings of the IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 24–28 July 2011, p. 60392282011.
42. Oksuz, I.; Ugurlu, U. Neural network based model comparison for intraday electricity price forecasting. *Energies* **2019**, *12*, 4557.

43. Jana, R.K.; Ghosh, I.; Sanyal, M.K. A granular deep learning approach for predicting energy consumption. *Appl. Soft Comput. J.* **2020**, *89*, 106091.
44. Naik, N.; Mohan, B.R. Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique. In *2nd International Conference on Emerging Technologies in Computer Engineering: Microservices in Big Data Analytics, ICETCE 2019*; Springer: Singapore, 2019, Volume 985, pp. 261–268.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).